

No	質問	回答
1	テキストクラスタリングの左のキーワードを一度に下位の言葉まで開く方法は無いのでしょうか。また、開いたキーワードを一括して階層も含めてExcelなどにコピー & ペーストする方法はありますか？ テキストクラスタリングでできた階層構造のデータをエクセルとかに出力することはできますか？	下位のワードの展開には、クリックが必要です（一括で開くことはできません）。展開後にキーワードをコピーアンドペーストしてExcel等での利用が可能になります。また、このとき階層構造を含めた形での出力はできません。 Excel等ファイル形式でのエクスポートはできません。
2	“ダミー公報”そのものは、リアル公報集合の中にどうやって入れるのですか？	配布資料の参考資料2(p.31～)で新たに説明を用意させて頂きました。
3	よくみるサーモマップでは、“島”になったり、大陸にくっついて“出っ張った半島”になるのを見ますが、2つの違い（内部ロジックなど）について、可能な範囲で教えて頂けないでしょうか？ 離れている距離は同じくらいでも、島になったり、半島になったり、ありますが…	該当するキーワードの頻度が多い場合にピークが高くなります。そして、それぞれのレコード（点）の距離は、キーワードの類似度が高いと近く、低いと遠くに配置されます。その結果として、全体として等高線が引かれ、島や半島に見えます。 例えば、半島に見える場合は、山頂間で共通するキーワードが存在するため、陸としてつながっております。一方で、島に見える場合は、近隣の大陸と共通するキーワードが比較的小さいことを示しております。
4	“山の頂点”にあるキーワードは、その周辺に集まる特許の中で、“一番頻出しているキーワード”という理解で合っていますか？	概ねその理解で良いです。 もう少し説明すると、単語/フレーズからトピックを選択するのに、TFIDF(Term Frequency/Inverse Document Frequency)法を用いて順位づけが行われております。当該単語/フレーズの文書中での頻度が高いとスコアは高くなり、さらに、当該単語/フレーズが他の文書には出現していない場合（よりユニークな単語）である場合にはスコアが高くなります。これにより頻度が高くても一般的な用語はトピックから除外されます。 そして、その山のまわりのレコードを最もよく表しているトピックが、“山の頂点”にラベルとして表示されます。
5	フレーズの切出しはどのようにしているのでしょうか	Derwent Innovationの内部で、語幹処理、類義語処理などを含む一定のルールを用いて単語及びフレーズの切り出し（グループ化）を行っております。
6	ストップワードの設定の仕方を教えて下さい	配付資料の別添資料 1 (p.22～)で、操作方法を新たに説明を用意させて頂きました。
7	特許公報中の表、図からもテキストは抽出されるのでしょうか	テキストは、明細書中の表や図そのものからは抽出されません（イメージからの抽出はされない）。なお、Derwent Innovationで扱えるテキスト形式のフィールドを解析対象とすることができ、「図面の説明」は明細書の一部として解析対象に含めることは可能です。
8	他社製品などでは、請求項記載の“パラメータ”を対象にしたアプローチができるものもおおくあるようですが、Derwent Innovationのテキストマイニングでは、そういった機能、か、手法がありますか？	パラメータの解析はできません。
9	今見せていただいているマップ(332件)は、検索集合（DI検索での）からテーマスケープ描かせたもの？それともテキストマイニング？で集めた母集合で、描かせたもの？でしょうか？	後者となります。DEMOでは①DI検索での母集団から、まずDWPI抄録-用途を解析対象とした用途マッピングを行いました。②その後、テキストマイニングの結果から特定のキーワードの領域を選択「切り出し」(332件)をして、それを母集団としてDWPI抄録-優位性を解析対象とした課題マッピングを行っています。
10	ノイズ除去の方法でストップワードを挙げられていましたがどの程度までノイズを除去されますか。	解析する母集団や解析対象により条件が異なり、一概には言えませんが、トライアンドエラーで進めていくことで、ノイズを除去していくことが可能です。
11	ノイズ除去の方法としてはストップワード以外に方法はありますか。	紹介させて頂いた「切り出し」(p.15～)はノイズ除去も可能です。また、「マップのオプション設定」-「設定トピック頻度（下限のしきい値）」を適切に設定することで、ノイズの除去ができる場合もあります。
12	テーマスケープの山の高さは、出願数をはじめ変更できるのでしょうか。	できません。マップ全体の高さや位置は母集団及び解析対象のテキストに寄与し、自ら編集を行うことはできません。